PB87-212882

Development of an Indoor Air Pollution Source Emissions Database

North Carolina Univ. at Chapel Hill

Prepared for

Environmental Protection Agency
Research Triangle Park, NC

Jul 87

U.S. Department of Commerce
National Technical Information Service
NTIS

PB87-212882

EPA/600/D-87/221
July 1987

DEVELOPMENT OF AN INDOOR AIR POLLUTION SOURCE EMISSIONS DATABASE

By

Parker C. Reist and James C. Reaves
University of North Carolina
Chapel Hill, NC 27514

James B. White
Air and Energy Engineering Research Laboratory
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

CR813460

EPA Project Officer
James B. White

AIR AND ENERGY ENGINEERING RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NC 27711

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL
INFORMATION SERVICE
SPRINGFIELD, VA 22161

TECHNICAL REPORT DATA (Please read Instructions on the reverse before completing)		
1. REPORT NO. EPA/600/D-87/221	2.	3. RECIPIENT'S ACCESSION NO. PB87 212882/AS
4. TITLE AND SUBTITLE Development of an Indoor Air Pollution Source Emissions Database		5. REPORT DATE July 1987
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) P. C. Reist and J. C. Reaves (UNC-CH), and J. B. White (EPA/AEERL)		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of North Carolina School of Public Health Chapel Hill, North Carolina 27514		10. PROGRAM ELEMENT NO.
		11. CONTRACT/GRANT NO. CR813460-01-0
12. SPONSORING AGENCY NAME AND ADDRESS EPA, Office of Research and Development Air and Energy Engineering Research Laboratory Research Triangle Park, NC 27711		13. TYPE OF REPORT AND PERIOD COVERED Published Paper; 5/87
		14. SPONSORING AGENCY CODE EPA/600/13
15. SUPPLEMENTARY NOTES AEERL project officer is James B. White, Mail Drop 54, 919/541-1189.		

16. ABSTRACT The paper discusses the design, structure, and theory of a microcomputer-based relational database which has been created to archive and retrieve published information concerning sources of indoor air pollutants. The database is designed to be used by researchers, architects, policy makers, and others as a research aid to provide indoor air pollution source data for estimating emissions or predicting health effects.

17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Pollution	Pollution Control	13B
Emission	Stationary Sources	14G
Data Processing	Indoor Air	09B
Atmosphere Contamination Control	Data Bases	
	Health Effects	06K
Environmental Biology		06F

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
Release to Public	Unclassified	8
	20. SECURITY CLASS (This page) Unclassified	22. PRICE

EPA Form 2220-1 (9-73)

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

AEERL-P-315
5R 6/22/87

DEVELOPMENT OF AN INDOOR AIR POLLUTION SOURCE EMISSIONS DATABASE

Parker C. Reist and James C. Reaves
Department of Environmental Sciences and Engineering
School of Public Health
University of North Carolina, Chapel Hill, NC 27514 USA

James B. White
Air and Energy Engineering Research Laboratory
U.S. Environmental Protection Agency, MD-54
Research Triangle Park, NC 27711 USA

ABSTRACT

The paper discusses the design, structure, and theory of a microcomputer-based relational database which has been created to archive and retrieve published information concerning sources of indoor air pollutants. The database is designed to be used by researchers, architects, policy makers, and others as a research aid to provide indoor air pollution source data for estimating emissions or predicting health effects.

I. INTRODUCTION

The study of indoor air pollution is an entirely new field, with most experimentation having occurred within the last six years. As a result, experimental results on indoor pollution sources may not be readily available to researchers, policy makers, or the general public. It therefore seemed a unique opportunity to make these data more available and to provide a framework for accumulating and disseminating new indoor air pollution source data as they are developed.

In a previous publication(1) the concept of an Indoor Air Source Emissions (IASE) database was explored and a "first draft" computer program presented for using this database. This program was written in *dBase III*[R] and was intended for use with IBM[R] and compatible personal computers. Although demonstrating an approach to finding and utilizing the broadly based literature on indoor air emission sources via personal computer, there was still much to be done to increase program speed and versatility. This paper reports on our progress toward these goals.

II. DESIGN GOALS

The following major goals were decided upon for design parameters of the IASE database system:

1) It needed to be useful to users with a wide range of interests, including researchers, architects, policy makers, manufacturers, builders, designers, and modelers.
2) It was desirable to make the system as widely accessible as possible. This has been a major consideration in determining the best combination of relational database management system (DBMS) and computer hardware to accomplish this goal.
3) It was necessary for the IASE system to satisfy a broad range of needs. For example, some users may be interested only in bibliographic information, whereas others may wish to use the data to analyze specific problems relative to the emissions data.
4) The IASE system must be "user friendly"; *i.e.*, uncomplicated to use. This makes the IASE database accessible to persons with little or no training in the use of computers.
5) Since it is anticipated that the amount of emission rate data will grow rapidly in the near future, the data had to be easy to update and modify.

III. PROGRAM DEVELOPMENT

The IASE project is broken down into two main parts: 1) development and updating of the indoor air pollution source database; and 2) development of the computer program to efficiently search the database and present items of interest to the user. These two main parts will be discussed separately.

Updating of the Indoor Air Pollution Source Database

On a previous project, EPA/AEERL's Indoor Air Branch began compiling indoor air source emissions data published from 1979 through 1984. Information published in the years 1985-1987 is now being added to the database. The types of information that have been made available are detailed in Table 1. The near exponential growth of indoor air data in the period 1985 through 1987 makes it necessary to plan for improved means of information transfer to the database.

Techniques have been developed to more efficiently search the literature and enter the appropriate data into the data set. Even so, the technical nature of the database and raw information is such that some analysis and interpretation is generally needed before entry. We have attempted to minimize this requirement. In addition, we expect to continually update the data set as new information becomes available.

Computer Program Development

We have chosen the commercially available program *dBase III Plus*[R] as the primary programming vehicle. However, because commercial DBMS programs must be as versatile as possible in performing many different tasks, but must still operate within the restriction of the available memory of the small computer, programs such as *dBase III Plus* may not give optimum performance for any particular task. We are therefore creating a hybrid system which incorporates assembly language sub-routines within the IASE structure.

This hybrid system approach offers real potential for increased speed and data storage capabilities. Although the hybrid system will use *dBASE III Plus* for its primary structure, assembly language sub-routines tailored to the specific needs of the IASE database will be used when they prove more efficient than the more broadly oriented routines of the commercial DBMS.

Additionally, the program is being made available in a format that will run directly from *PC-DOS*[R] or *MS-DOS*[R] operating systems. Possession of *dBASE III Plus* or any other programming language is not required to use the IASE system.

IV. DESIGN PHILOSOPHY

General

The Indoor Air Source Emission database will provide the user with the ability to extract pertinent information from available research literature concerning sources of indoor air pollutants.

To operate the program the user is neither required to have any previous knowledge of the information held in the program data files nor knowledge of the structure of the files. Although the EPA/UNC group plans to maintain the overall database and add new information as it becomes available, users will be able to add to or modify their own copies to suit their individual needs and data requirements. For example, if it is found that data in one particular area are superfluous to the user's needs, they can be deleted. On the other hand, users will be encouraged to add new data entries to their own copies, and to forward these data to EPA/UNC for inclusion in the main database.

In order to be available to the widest audience possible, the program runs on IBM or IBM compatible personal computers minimally configured with two floppy-disk drives. We do, however, recommend a hard disk and an 80286 microprocessor. In the present version a complex search on three identifiers takes up to 2 minutes on an 8088 microprocessor equipped IBM PC, while an 80286 equipped machine performs the same search in 17 seconds. As the database grows, fast retrieval speeds will become increasingly important. By increasing data retrieval speed, the user has more time to use available sub-programs concerned with data analysis or predictions of source emissions rates.

Program Design

IASE is divided into three main segments: Data Query, Data Input, and Data Output.

Data Query is the main interactive section of the IASE system. Its function is to filter the amount of information available from the main database so that only information of interest to the user is presented. This process is accomplished through a series of menu screens from which the user makes selections and assigns logical relationships.

Areas within the database where the user may choose specific topics of interest are: pollutant name, source material, source class, author name, date of publication, and journal of publication. Up to three identifiers may be chosen at one time.

Program search algorithms and memory techniques have also been considered in the program design. When the user has selected topics of interest, the program evaluates the selections and performs three functions via a rudimentary artificial intelligence system:

1) Automatic assignment of boolean operators (and, or, nor) according to the most probable relationships between the chosen identifiers.
2) Grouping of the specified search parameters into blocks allowing the most efficient search of the database.
3) Review of a memory file to determine if any of these blocks have been requested by any previous user.

If a record of a previous search is found, that record is extracted from memory and the search is not repeated. If a record is not found, a search is performed and the record is stored in memory. One advantage here is that the program automatically tailors itself to the needs of the user. Frequent usage in a particular area permits building of a stored memory for searches within that area. This use of a memory file has two advantages over more traditional database search algorithms; it results in greatly decreased access time to data, and allows much larger databases to be used without much of the usual corresponding increase in search time.

For each block of grouped search parameters, memory records are found or created containing the location of information pertaining to that block's search parameters. Once memory records exist for all blocks, the blocks themselves are compared. The results of these comparisons are the physical locations within the database of information associated with the more complex search parameters specified by the user.

The program's ability to modify its own search procedures by grouping search parameters and by assigning boolean operators meets the definition of an artificial intelligence system. Even this rudimentary artificial intelligence capability significantly improves both the performance of the system and the simplicity of the user interface.

Data Input, or data entry, is critical to the correct operation of any relational data retrieval system. For example, punctuation, spelling and consistency of entry, and even the case of the characters entered

may mean the difference between accurate retrieval or effective loss of some information.

Data input for this database is particularly complex. For example, much of the information entered into the database comes from published articles listing multiple pollutants, with each pollutant possibly having multiple test conditions. This requires data entry for each pollutant/test condition combination. Since one of the main requirements of this database system is the capability for the user to input his or her own data, the data entry portion of the program must be straightforward and call for as little repetitive typing as possible.

Data is entered using self-explanatory data entry screens similar to menu screens, with the user being prompted for those entries where errors could be significant. In addition, error checking is used whenever possible to test the validity of the input data. Context-sensitive help screens are available as an additional method to minimize errors upon data entry.

Data Output for the IASE system is just as complex as data entry. This presents a problem in that all of these data may not be required by the user,and if they are,they will require several screens or printed pages per individual record. To solve this problem the user is presented with a list of records that have matched the input criteria, with only a minimum of identifying information included. The user may then pick individual records to view in detail.

V. PROGRAM AVAILABILITY

Information concerning availability of developed draft IASE programs is available from James B. White.

VI. REFERENCES

1. Crum, Jane M.,"Indoor Air Pollution Source Database, "Presented at 79th Annual APCA, Minneapolis, MN, June 22-27, 1986 (EPA-600/D-86-082:NTISPB86-191376).

TABLE I
Currently Defined Fields

TEST CONDITION
Conditioning
Duration
Test description
Operating parameters
Temperature
Humidity
Ventilation rate
Size of material
Loading
Fuel rate
Fuel composition
Fuel value
Other conditions

BIBLIOGRAPHY
Author
Journal
Title
Publication date
Abstract
Test facility category
Test facility description

SAMPLE DESCRIPTION
Sample type
Sample rate
Sample duration
Sampling method
Sample size
Analytical method

POLLUTANT
Pollutant
Comments
Measurement
Calculated emissions

SOURCE
Category
Class
Type
Description